



Early science learning with a virtual tutor through multimedia explanations and feedback on spoken questions

Jarkko Hautala¹ · Doris Luft Baker² · Aleksi Keurulainen⁶ ·
Miia Ronimus³ · Ulla Richardson⁴ · Ronald Cole⁵

© Association for Educational Communications and Technology 2018

Abstract The purpose of this pilot study with a within-subject design was to gain a deeper understanding about the promise and restrictions of a virtual tutoring system designed to teach science to first grade students in Finland. Participants were 61 students who received six tutoring science sessions of approximately 20 min each. Sessions consisted of a sequence of narrated multimedia science presentations during which a virtual tutor explained science phenomena displayed in pictures. Narrated science explanations were followed by one or more multiple choice questions with immediate feedback about

✉ Jarkko Hautala
jarkko.v.hautala@jyu.fi

Doris Luft Baker
dluftdebaker@mail.smu.edu

Aleksi Keurulainen
lxi@iki.fi

Miia Ronimus
miia.ronimus@nmi.fi

Ulla Richardson
ulla.a.richardson@jyu.fi

Ronald Cole
rcole@boulderlearning.com

¹ Center for Applied Language Studies, Department of Psychology, University of Jyväskylä, Mattilanniemi 6, Jyväskylä, Finland

² Department of Teaching & Learning, Southern Methodist University, 6401 Airline Rd, Dallas, TX 75205, USA

³ Niilo Mäki Institute, Asemakatu 4, 40014 Jyväskylä, Finland

⁴ Center for Applied Language Studies, University of Jyväskylä, Mattilanniemi 2, 40014 Jyväskylä, Finland

⁵ Boulder Learning Inc., 2960, Center Green Court, Suite 200, Boulder, CO 80301-5406, USA

⁶ Qvantel Ltd., Kauppakatu 28, 40100 Jyväskylä, Finland

students' choices and a possible second attempt, during which students reached 97% accuracy. A pretest and posttest was administered to assess students' ability to reason about the science and to transfer knowledge to new contexts. Results indicated significantly greater improvement in the understanding of the science concepts taught during the tutoring sessions, relative to the concepts that were not taught. Results from the surveys administered to teachers and students indicated that the program was well received. Detailed analysis of student error responses provided a deeper understanding about the complex interplay between students' prior knowledge, the way topics were taught in the multimedia lessons, and the way learning was assessed. Findings from the quantitative and qualitative analyses are discussed in the context of designing high quality lessons delivered through a virtual tutoring system.

Introduction

Evidence from national and international assessments has indicated that a significant proportion of students worldwide fail to meet grade-level academic standards in science (National Center for Education Statistics [NCES] 2011; Programme for International Student Assessment [PISA] 2012). For example, in the United States, over 60% of fourth-grade students were rated as “not proficient” in science (NCES 2011). Although in Finland, student science outcomes on international assessments are higher, performance has declined in the last decade (Kupari et al. 2012; PISA 2012). Given that students' attitudes and motivation to learn science develop before the age of 14 (Osborne and Dillon 2008), it is critical to engage early-grade students in effective science instruction across the globe.

Science learning can be seen as a process of acquiring progressively new core ideas, concepts and principles to explain natural phenomena observed (National Research Council [NRC] 2012). In the present study a virtual tutoring application [i.e., a lifelike interactive computer character providing guided tutoring in immersive multimodal environments (Wise et al. 2005)] teaches first grade student new ideas and principles to explain various animal behaviors, structures and functions, by means of short audio-visual lessons interleaved by multiple choice problems with guided feedback. The lessons are organized into learning progressions, starting with an intriguing question (e.g. “What is an insect?”), then teaching required concepts one by one with exercises (body parts of an insect), and finally practicing the aim of the lesson (identifying insects from other resembling creatures).

Previous research has indicated that virtual tutors can provide high-quality, individualized, and highly engaging science teaching equivalent to human tutoring (Cohen's $d = 0.74$) in adult students (Lieberman et al. 2009; Van Lehn 2006, 2011; Wise et al. 2005). Results among children have also been encouraging. Ward and colleagues (2011, 2013) examined the science learning of third- to fifth-grade students who received 16 supplementary 15-min tutoring sessions by expert humans or spoken multimedia dialogues with virtual tutors relative to the science learning of students who did not receive tutoring as a supplement to classroom instruction. Statistically equivalent results were obtained for human and virtual tutors ($d = 0.62$ for the human tutors and $d = 0.56$ for the virtual tutor). Dalacosta et al. (2009) compared the effectiveness of human-delivered instruction and cartoon-style multimedia lessons (including multiple-choice questions [MCQs] with corrective feedback) in teaching the concepts of mass, volume, and density to fifth-grade students. Their findings indicated that the virtual-tutor group outperformed the group receiving human-delivered instruction.

Moreover, in general, reviews have also indicated that having a visually present animated tutor does not seem to produce any learning gains over having animated tutors that are not visually present (Heidig and Clarebout 2011; Schroeder and Adesope 2014). However, this hypothesis has not been tested among young children. In a preliminary assessment of young children's impressions of studying with animated tutors, kindergarten and first-grade students reported that the virtual tutor they saw was smart, cared about them, and acted like a real teacher who helped them learn to read (Cole et al. 2007). In the studies of Ward et al. (2011, 2013), over 75% of students reported that they were more motivated to study science after working with the virtual tutors.

In sum, recent studies have suggested a great potential for well-designed virtual tutoring systems to engage and motivate primary school students to learn science and to achieve learning gains comparable to those from human tutoring. However, to our knowledge, the applicability of virtual tutors is yet to be studied in children who have just started their school path and are still learning basic scholastic skills such as reading and comprehension in content areas such as science.

Theoretical framework for the design of the virtual tutoring System

The design of the virtual tutoring system used in this study, Mindstars Books (MSB), integrates ideas from cognitive theory of multimedia learning (Mayer 2014), formative assessments (Black and Wiliam 1998), and the Dual Situated Learning Model (She and Liao 2010). While these theories provide a general framework for the design of the presentation format, instructional interaction, and science content progressions, respectively, special attention is paid on how to design the system to optimize the learning in small children.

The cognitive theory of multimedia learning states that narrated multimedia presentations help learners construct rich multimodal mental representations that lead to the deep learning of concepts (Mayer 2014). A large body of research has indicated that relative to other presentation modes such as texts with pictures, well-designed narrated multimedia presentations, in which a spoken voice explains concepts presented in illustrations or animations, optimize both the short-term retention of information and the transfer of learning to new tasks (Mayer 2014). Additionally, a meta-analysis by Heidig and Clarebout (2011) suggested that spoken explanations combined with visual illustrations, simple presentations in small steps (Adesope and Nesbit 2012; Sweller 1994), and control over the pacing of lesson content improve learning in multimedia settings. These design principles minimize the cognitive resources required for using the application, allowing users to maximize their focus on acquiring content knowledge.

Early-grade students are only beginning to develop their basic scholastic and cognitive skills such as reading and listening comprehension. However, the early levels of these skills predict later school achievement from kindergarten to second grade (La Paro and Pianta 2000). Similarly, students of this age do not yet possess the metacognitive skills required for self-regulated learning approaches (Dignath and Büttner 2008). Thus, to engage all students—irrespective of their scholastic skill level and cognitive capabilities—in science learning, educational software targeted to small children needs to provide a clear structure and guidance for the learning activities (Dignath and Büttner 2008). It should also avoid using features that require practiced skills and knowledge such as reading (Wang et al. 2010).

The MSB design was also informed by the use of formative assessments where the learning goal and the learning progression are made explicit to the learner, and the learning occurs in a dialogue with a more knowledgeable person such as the teacher (or in this case, a virtual tutor; Black and Wiliam 1998). Multimedia learning systems can interact with a learner by providing individualized instruction such as MCQs with explanatory and corrective feedback (Black and Wiliam 1998; Kintsch 2005), where students have an opportunity to reconsider their answer and make another attempt at the task (Craig et al. 2000; Lin et al. 2013; Yoshida 2008).

Most current science learning models (Duit and Treagust 2003; Hong and Diamond 2012; She and Liao 2010) follow the principles of formative assessment. For example, the Dual Situated Learning Model (DSLML) by She and Liao (2010) facilitates science learning by (a) engaging and motivating students to understand a particular phenomenon, (b) taking into account students' prior knowledge for them to make sense of the information presented, (c) scaffolding the lesson progression and introducing information in small conceptual chunks, and (d) presenting students with opportunities to test and apply their knowledge. Given that the learning of science concepts is not always intuitive and often requires fundamental changes in thinking, many science learning models, including DSLML, stress the importance of producing and resolving cognitive conflict within the learner's mind (e.g., Duit and Treagust 2003; She and Liao 2010). However, research has demonstrated that cognitive conflict may actually facilitate learning only in students with higher capabilities for logical thinking, but not in younger students or students with less capability for logical thinking (Kang et al. 2004). Young students may experience cognitive conflicts for various reasons, including common misconceptions about science but also due to very limited previous knowledge of the topic, which may lead to perceptual levels of confusion when reading science text. Research indicates that clarifying the common misconceptions appears to be more effective for students with limited background knowledge than explicitly creating cognitive conflicts that students have to resolve (Smith et al. 1994). Thus, in this study, we attempted to answer the following research questions:

1. *Can first-grade students learn science through a virtual tutoring system such as the MSB?*
We investigated this question by analyzing (a) the immediate understanding of concepts presented in the MSB; (b) the long-term learning gains on science assessment presented before and after the entire 6-week course; and (c) whether the MSB had a differential effect on pretest–posttest science-assessment gains based on prior knowledge and students' individual skills in reading and listening comprehension.
2. *Are the MSB a useful and feasible educational tool in authentic educational settings?*
We investigated this question through surveys administered to both students and teachers. In addition to the learning benefits, educational technology should provide time and cost benefits and be liked and valued by students and teachers.
3. *How does the visual presence of an animated virtual tutor affect students' learning and/or enjoyment of the program in comparison to just hearing the tutor's voice?*
We investigated this question by presenting half of the MSB to each student with the tutor's face on screen and the other half with the tutor's face off screen (i.e., where the students only heard the tutor's voice).
4. *What design features might have affected students' knowledge acquisition?*
To identify the students' most common misconceptions and other possible obstacles to learning, we conducted a detailed item-specific analyses of MSB MCQ and post-test science assessment questions.

Method

Research design

This pilot study used a quasi-experimental within-subject design. Students were assessed at pretest (i.e., before exposure to the MSB) and at posttest (i.e., after exposure to the MSB). Pretest scores were taken into account when calculating the effect of the intervention on student outcomes.

To answer the third research question, students were randomly presented with three MSBs with the tutor's face visible (face on and voice on) or not visible (face off and voice on). The order of presenting the face on and face off versions of the MSBs was counterbalanced across participants; none of the students received all three face-on or face-off versions in succession. The order of the lessons was the same for all students, progressing from conceptually simpler lessons to more complex ones.

Participants

Sixty-three first-grade students (28 males and 35 females) from three classrooms in an elementary school from a town in central Finland participated in the study. The school was a teacher training school associated with a nearby university. Participants were 6- and 7-year-old students who spoke Finnish as their native language. Two of the students were excluded from the analysis because they were absent from school during most of the sessions, so the total number of students in the final analysis was 61.

Materials

The MSB design followed the theoretical framework explained earlier. Figure 1 presents the MSB user interface, including a virtual tutor giving spoken explanations with animated mouth movements, a content window typically showing highlighted pictures timed with spoken explanations, and a self-pacing button.

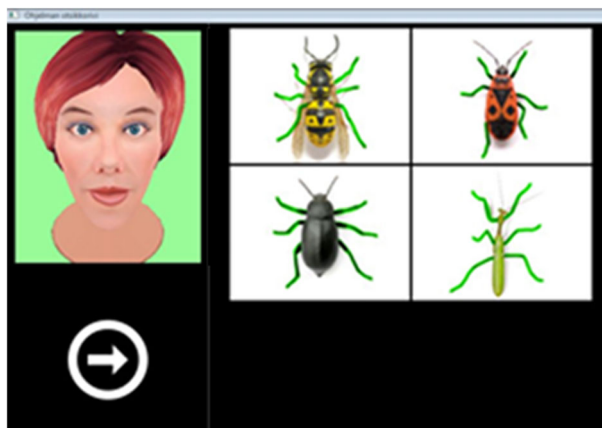


Fig. 1 A screen capture of Mindstars Books

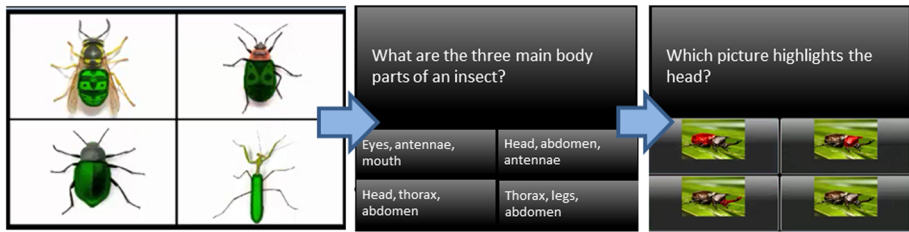


Fig. 2 Example of a lesson sequence. On the first picture, the tutor said, “Every insect has three main body parts. What are these three main body parts? The three are the head, the thorax, and the abdomen.” On the second picture, the tutor instructed, “Listen carefully and select the best answer. What are the three main body parts of an insect?” On the first click, the tutor spoke aloud the written answer choice, and on the second click, the selection was made. The correct answer was reinforced by prompts such as “That’s right! Every insect has a head, a thorax, and an abdomen. All adult insects have three main body parts!” Wrong answers such as “Thorax, legs, abdomen” were followed by a corrective hint such as “Legs are not a main body part. Legs are attached to the thorax. The thorax is one of the main body parts, as well as the abdomen. You are missing one main body part. What is it?”

Figure 2 illustrates a sequence of verbal and pictorial multimedia science explanations ending with an MCQ with formative feedback. At logical stopping points, MCQs were presented to assess students’ understanding of the vocabulary and science they had just learned, with immediate feedback contingent upon their answer choices. The questions were read aloud to students; students could click on printed answer choices to hear the choices read aloud again by the tutor before selecting an answer. The answer choices to some of the questions consisted of pictures. A correct first attempt was followed by positive reinforcement (“Good thinking”) and an expansion of the correct answer by the tutor. An incorrect first attempt was followed by a hint and a second choice (see Fig. 2). If the second attempt was incorrect (which was rare), the question was repeated, and the correct answer was presented along with its explanation. Throughout the lesson, an option to repeat explanations, questions, and answer choices supported students’ comprehension.

The intervention in this study consisted of six MSB that focused on teaching life science: (a) *Five Senses*, (b) *How Do Animals Move?* (c) *What Do Animals Need to Live?* (d) *How Are Animals Covered?* (e) *What Is an Insect?* and (f) *The Life Cycle of a Butterfly*. The MSBs were originally developed in English by our collaborators in the United States. The content of the MSBs were aligned with the Next Generation Science Standards on the Disciplinary Core Idea of Structure and Function of animals (NRC 2012). Finnish translations of the MSBs were created, and life science terms and concepts were reviewed by a university lecturer in early biology education. A female Finnish speaker recorded the virtual tutor’s speech. The tutor’s mouth movements while speaking Finnish was synchronized with the recordings of the female voice.

In line with the DSLM framework, all MSBs began with an introduction to the topic and the scientific problem. For example, the lesson *What Is an Insect?* first posed the question “How do we know if an animal is an insect?” along with pictures illustrating the great diversity in insects’ appearance. To activate background knowledge, we selected familiar and elementary animal-related topics such as “How are animals covered?” for the lessons, and we used accessible language and vocabulary. In addition, the content was designed to resolve possible cognitive conflicts stemming from typical misconceptions by addressing such issues in lessons and through the answer choices of MCQs and their associated feedback. For example, in the insect lesson, the common misconception that spiders are insects was addressed by explicitly teaching the characteristics of an insect (i.e., an insect

has three body parts, six legs attached to the thorax, and two antennae attached to the head) versus the characteristics of a spider (i.e., a spider has two body parts, eight legs, and no antennae). After the core concepts and vocabulary were taught, they were then summarized and integrated. In this phase, the insect lesson demonstrated why a damselfly is an insect while a spider is not an insect. Finally, the lesson ended with several MCQs touching on the original goal of the lesson (e.g., asking the student to identify an insect from other similar creatures such as a centipede or a scorpion), thus reducing common misconceptions related to animal classifications, for example.

Measures

MSB MCQ responses

For enabling the analysis of student's immediate science understanding as reflected by their selections on multiple choice questions presented during studying the MSBs, the system stored all the user behaviors with time stamps in a log file. The students' choices were used to calculate accuracy rates and analyze incorrect responses.

Pretest and posttest science assessment

Long-term retention and transfer of learning based on information presented in the MSB

A total of 24 researcher-developed MCQs were asked in the pretest and posttest (Table 3 in Appendix 1). These questions were directly related to the content taught in the MSB. They were designed to assess students' deep understanding of content taught in the six MSB rather than simply their recall of facts. For example, the need for oxygen was illustrated in the MSB by a turtle whose nose was above the water's surface, with the narration explaining that turtles need to come to the surface to breathe air. In the pretest and posttest, students' understanding of this science concept was measured by the question "Which picture represents the need for oxygen?" (Fig. 3). The correct answer was the picture showing a porpoise submerged in water and exhaling bubbles near the surface. Selecting this picture required the student to reason that the porpoise was breathing out under water but would need to breathe in once it came to the water's surface.

Questions about first-grade science not taught in the MSB

In addition to the questions about science taught in the MSB, we also asked 15 researcher-developed science questions that targeted content that was not taught in the MSB but was relevant to first-grade science instruction (Table 3 in Appendix 1). These questions were designed to evaluate how much science knowledge can be learned as a result of repeated testing and transfer from other learned knowledge (Adair et al. 1989). To ensure the age appropriateness of the questions, the themes were taken from the lessons in a first-grade life science textbook to which students had not yet been exposed. An example of a control question is "Which body system distributes oxygen?" The correct answer was a picture representing the circulatory system.

The order of the 39 questions (i.e., 24 on the taught content and 15 on the not-taught content) was different for the pretest and posttest. Each pretest–posttest question had three

Select a picture representing a need for

- | | | | | |
|------------|----------------------------|----------------------------|----------------------------|----------------------------|
| a) water | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | 4 <input type="checkbox"/> |
| b) shelter | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | 4 <input type="checkbox"/> |
| d) oxygen | 1 <input type="checkbox"/> | 2 <input type="checkbox"/> | 3 <input type="checkbox"/> | 4 <input type="checkbox"/> |



Fig. 3 An example of pretest and posttest multiple-choice questions

to seven alternative answers presented as pictures; students had to choose the correct answer or answers (Fig. 3). Two of the questions were open ended and required a one-word written answer. Cronbach's alpha for the test, calculated from the pretest scores, was 0.716. Figure 3 provides an example of the material presented to students. Table 3 in Appendix 1 provides a list of all the questions.

Lesson evaluations

To assess students' opinions about the usability and likability of each MSB, as well as their learning experiences with the MSB, we asked them to answer a paper questionnaire composed of 10 MCQs after each study session. The questions are provided in the Table 2 at the Results -section.

Evaluation of the intervention

To assess students' opinions of the MSB experience, we asked them to answer a user experience questionnaire with seven MCQs (e.g., "Which version of the MSBs do you prefer?" [Talking head, No talking head]) and two open-ended questions at posttest (Appendix 3).

Reading skills

To evaluate whether students' reading skills contributed to their learning gains, teachers administered a standardized test of word-reading skills (ALLU TL2, version B; Lindeman 1998) as a group test in the fall prior to the study. In this 80-item paper-and-pencil test, each student matched four printed words to four pictures by drawing a line between each

word and one picture. The score was the number of correct answers given within a two-minute time limit.

Listening comprehension

To evaluate whether students' listening comprehension skills contributed to their learning gains, teachers administered a standardized test of listening comprehension skills for first-grade students (ALLU KY; Lindeman 1998) as a group test during the fall prior to the study. In this test, the teacher read a story aloud twice, after which participants' comprehension was assessed using six comprehension questions read aloud. To answer the questions, students selected one of four alternatives on their answer sheets; there was no time limit. The scores ranged from 0 to 2 points depending on the selected alternative, resulting in a maximum score of 12 points.

Teachers' survey

After the MSB intervention, all teachers responded to an e-mail survey composed of seven open-ended questions (e.g., "According to your experiences, how useful is the MSB type of educational technology in an early education context?") about the feasibility of using the MSB and this type of new learning technology in general as a supplement to their classroom science instruction (Appendix 3).

Procedure

The study was conducted from January to March 2014 in three first-grade classrooms. A fifth-year university student of teacher education administered the intervention and the assessments. After the pretest, the students studied one MSB in a single session in the school's 12-seat computer lab each week. During a single 45-min lesson period, half of the class visited the lab for 20 min, followed by the other half of the class. The students wore headphones and used a mouse to hear and repeat utterances, turn the pages of the lesson, listen to and repeat questions and answer choices, and select their answers. Students were assigned user identification numbers (to preserve anonymity), which the research assistant entered for each student prior to each study session. Posttest was administrated on next week after the last MSB study session.

Pretests, posttests, and student questionnaires were administered as a group test in the classroom. Instructions for answering were given prior to each test, and all the questions were first read aloud to the students. A projector was also used to display answer choices on a screen. Students wrote their answers on a scoring sheet. When necessary, the research assistant or the teacher helped students enter their intended written responses to the open-ended questions. The pretest and posttest sessions took about 45 min each.

Results

Can first-grade students learn science while using the MSB?

Within-MSB science understanding

While studying the MSB, students answered a total of 3409 MCQs. The percentage of correct answers after students' first attempt was 89.2%. Of the 368 questions answered incorrectly on the first attempt, 278 (i.e., 81.1%) were answered correctly on the second attempt. Thus, a total of 3319 of the 3409 questions (97.4%) were answered correctly on either the students' first or second attempt. These results indicate that (1) the questions were generally well aligned with the content taught in the lessons, (2) students were able to answer most of the questions correctly on their first attempt, and (3) the hints following incorrect first attempts were effective in scaffolding learning, given that over 80% of the second attempts were correct.

Pretest and posttest science assessment

To analyze learning effects, 10 questions on taught content and 10 questions on not-taught content were selected (marked as T and N, respectively, in Table 3 in Appendix 1). We included only questions that (a) were not at ceiling on the pretest (thus excluding four questions with $> 85\%$ accuracy), (b) had four answer choices with one unambiguous correct answer (15 questions excluded), and (c) featured a topic that was explicitly taught in the MSBs (in the case of questions on taught content). The excluded questions were originally included to evaluate whether students could reliably answer open-ended questions, questions with more than four choices, and questions with several possible correct answers.

Within-subject, repeated-measures analysis of variance (ANOVA) with two levels—the within-subject factor of measurement (pretest, posttest) and the within-subject factor of question type (content taught, content not-taught)—was used to assess the effect of the MSB on the students' science knowledge. Paired t-tests were used to further analyze significant interactions.

To study whether the MSB were producing meaningful science learning, we compared differences in pretest–posttest scores of questions on taught content and not-taught content (Fig. 4). A significant two-level interaction between measurement and question type, $F(1, 60) = 6.87$, $p = 0.011$, $\eta_p^2 = 0.103$, indicated larger learning gains on questions related to taught content relative to questions on non-taught content. There was a significant gain from pretest to posttest on questions related to taught content, $t(60) = -5.54$, $p < 0.001$, Cohen's $d = 0.71$, but not on questions related to not-taught content, $t(60) = -1.47$, $p = 0.148$, Cohen's $d = 0.19$. Moreover, at pretest, there were no significant differences between the different type of questions, $t(60) = -1.5$, $p = 0.127$, Cohen's $d = 0.19$. Additionally, there was a significant main effect of measurement, $F(1, 60) = 25.28$, $p < 0.001$, $\eta_p^2 = 0.296$, indicating overall higher scores in the posttest relative to the pretest. The main effect of question type was not significant ($F = 0.237$).

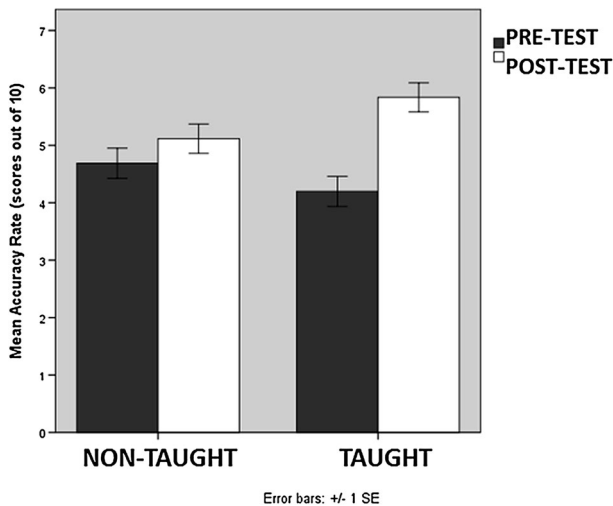


Fig. 4 Pretest and posttest scores on questions related to taught and non-taught content

Correlations between pretest–posttest learning gains, prior knowledge, and reading and listening comprehension

Given our small sample size and the insufficient statistical power to use the reading and listening comprehension measures as covariates, we only examined the strength of the relationship between reading and listening skills on learning gains through correlation analysis. We found a significant negative correlation ($r = -0.60$) between pretest scores and learning gains, as indicated in Table 1, suggesting that students with less prior knowledge of content taught in the MSB learned more than students with more prior knowledge. While this correlation is partially explained by the students with high prior knowledge reaching ceiling, a more important aspect is that the students with poor prior knowledge were also learning with the MSB. Interestingly, the level of reading and the level of listening comprehension skills were not associated with larger pre-to-posttest learning gains, which also indicates that basic scholastic skills suffice for learning science with the MSB. However, the accuracy ratio in the MCQs correlated positively with standardized listening comprehension ($r = 0.42$) and reading ($r = 0.25$), suggesting that students with better listening comprehension skills made fewer wrong selections when studying the MSB. Despite this, the second answer choice seemed to guarantee learning also among students who may have difficulties in listening comprehension.

Table 1 Correlations between reading skills, listening comprehension skills, learning gain, and prior knowledge in terms of pretest score on taught matters in Science Assessment

	MCQ	Reading	Listening	Gain
Reading	0.25 ¹			
Listening comprehension	0.42**	0.15		
Learning gain	− 0.02	− 0.16	0.14	
Science assessment	0.15	0.24	− 0.13	− 0.60**

¹ $p = 0.051$, * $p < 0.050$, ** $p < 0.001$

Are the MSB a useful and feasible educational tool in authentic educational settings?

Lesson evaluations

Table 2 shows students' responses to a set of questions immediately after using each MSB. Overall, 78% of the students reported that they liked the MSB a lot, 77% were eager to study the next lesson, 56% said they were more excited about science after using the MSB, and 74% felt the content was easy to learn. Slightly less than half of the students felt they had learned a lot from the MSB, although most of the students reported that they did not want to study the lesson a second time. We did not find any substantial or systematic differences in ratings between the different MSB.

Evaluation of the intervention

Students had a highly positive experience of the science course: 89% considered the MSBs good, 83% would study these kinds of materials in the future, 91% would study these kinds of materials at least once a week, 67% preferred to study in the computer class rather than their classroom, 65% preferred to study the MSBs at school, and 30% preferred to study these kind of materials at home. Finally, 35% reported liking science more than before, and 60% said they liked science as much as before.

Teacher survey

The teachers were generally impressed with the program based on their students' eagerness to study with the MSBs. They stated that their students, including those who had attention and self-regulation difficulties, performed very well with the MSB. The teachers were

Table 2 Students' responses to the usability survey data conducted after reading each of the six books

Question		%		%		%	Z
1 Could you hear the teacher well?	Yes	95.7	Between	3.1	Badly	1.1	– 0.446
2 Did the book work properly?	Yes	92.9	Don't know	6.8	No	0.3	– 0.637
3 Was the book easy to follow	Yes	92.6	Don't know	6.6	No	0.9	– 0.232
4 How much did you learn?	A lot	43.6	A little	37.9	Not at all	18.5	– 0.631
5 From my opinion the content was	Easy	74.4	Just right	23.9	Difficult	1.7	– 0.849
6 From my opinion, the teacher was?	Good	79.6	Between	18.1	Not very good	2.3	– 0.716
7 How much did you like the book?	A lot	77.6	Little	19.8	Not at all	2.6	– 1.33
8 Would you read it again?	Yes	59.3	Maybe	28.8	No	12.0	– 1.75
9 Are you looking forward to the next book?	Yes	76.9	Maybe	19.7	No	3.4	– 0.992
10 How excited are you about science?	More	56.0	Same	41.4	Less	2.6	– 0.446

The rightmost column shows the Wilcoxon rank test value for comparing answer distributions between the face-on and face-off versions of the books

interested in the content and the pedagogical strategies of the MSB, the degree to which the content encouraged students to think independently, and whether the MSB were effective in emphasizing the most important things to learn. They stressed that any software should be easy for students to use if the teachers are to ask their students to use it independently. They expressed a need for scientific evidence of the beneficial effects of educational softwares in the early elementary grades.

How does the visual presence of an animated virtual tutor affect student learning and/or enjoyment of the program in comparison to just hearing the agent's voice?

To study the effect of the visual presence of the virtual tutor on learning, a paired t test was used to compare learning gains in the science assessment between the face-on and face-off versions of the MSB, as well as MSB MCQ accuracies. There was no difference in pretest–posttest learning gains on questions related to taught content produced by the MSB in the face-on ($M = 0.90$, $SD = 1.40$) and face-off versions ($M = 0.74$, $SD = 1.81$), $t(60) = -0.567$, $p = 0.573$, Cohen's $d = 0.07$. Furthermore, there was no difference in the mean accuracy proportions of responding MSB MCQ questions between the face-on ($M = 0.896$, $SD = 0.09$) and face-off versions ($M = 0.889$, $SD = 0.11$), $t(60) = 0.639$, $p = 0.525$, Cohen's $d = 0.08$.

A non-parametric paired-sample Wilcoxon test was used to compare students' responses to lesson evaluation questions after studying the face-on and face-off versions. The results of the Wilcoxon rank test on responses to lesson evaluation questions did not differ between the face-on and face-off versions (rightmost column of Table 2).

We also used a binomial test to examine whether students preferred to use the face-on or face-off versions of the MSBs. This preference was measured by the question "Which version of the MSBs do you prefer?" [i.e., the one with the Talking Head, or the one without the Talking Head] in the Evaluation of the Intervention—query (Appendix 3). Twenty-six students chose the face-on option, and 34 students chose the face-off option. Binomial testing indicated that the responses were distributed equally between the two categories ($p = 0.366$).

What design features might have affected students' knowledge acquisition?

To identify the questions where learning did or did not occur, we ran a one-way non-parametric paired-sample Wilcoxon test for pretest versus posttest responses to each question (Table 3 in Appendix 1). To identify the students' common misconceptions especially in questions where learning did not occur, the frequencies of the most common errors for each question were derived from log files, and are being provided in Tables 3 and 4 in Appendices 1 and 2. The last columns of the appendices contain our possible interpretation of why some questions were answered incorrectly.

Some errors may represent perceptual confusions. For example, one posttest question asked the students to choose the animal that has scales. Several students answered "frog" instead of "fish." The frog had coloring resembling large scales, which may have misled some students.

Some errors were exceptions to a rule. For example, about 19% of students answered that animals get water "by eating" instead of "by drinking." This confusion was probably due to the lesson explicitly teaching that some animals get water from the food they eat.

Some errors may reflect deeper challenges in learning. The pretest and posttest questions about the lesson “*The Life Cycle of a Butterfly*”, which presented each phase of the life cycle of a butterfly (from egg to caterpillar to chrysalis to adult), were challenging for many first graders. They answered the related posttest questions with a mean accuracy of 41%.

Facts that contradicted prior conceptions also led to errors. For example, students believed that people have smooth skin, even though the lesson “*How Are Animals Covered?*” explained that humans are covered with hair, which is the same as fur. While the lesson taught that frogs have smooth skin without hair, when the students were asked which animal has smooth skin, 34% chose the picture of a human arm (with some hair on it) rather than the picture of a frog.

Students seemed to have difficulties in learning certain dynamic concepts (i.e., crawling and slithering) from looking at pictures of lizards (which crawl) and snakes (which slither).

Discussion

Did first-grade students learn science through the virtual tutor system?

The 89% accuracy in answering the MCQs presented while studying the MSBs indicates that students comprehended well what was presented and were able to reason about science concepts using the vocabulary terms that appeared in the narrated multimedia presentations. After a second attempt, the accuracy approached ceiling (97%), showing that the hints provided after incorrect answers successfully corrected the young students’ misconceptions (Black and Wiliam 1998; Kintsch 2005; Yoshida 2008).

The significant learning gains from the pretest to posttest on questions related to the topics that were taught in the MSBs expanded the initial learning to meaningful and long-term science learning; this result is consistent with those found among older students (Dalacosta et al. 2009; Gardner et al. 1992). Moreover, the fact that no learning occurred in questions on non-taught, age-appropriate science content provides further evidence that learning was not due to external factors such as repeated science assessment with the same items, which may induce learning in some cases.

Perhaps the most encouraging findings from this study were that students (1) with the lowest prior science knowledge showed learning gains in science assessment from pretest to posttest, and (2) that students’ academic reading and listening comprehension skills did not correlate with these learning gains. These findings are in line with our previous research on older, third- to fifth-grade students’ spoken dialogues with a virtual tutor, which demonstrated that the greatest learning gains were from students who scored the lowest on the pretests (Ward et al. 2013).

Altogether, these results are among the first to demonstrate the applicability of specifically designed virtual tutoring systems for teaching science in the earliest grades. Based on these findings, it is reasonable to assume that even young children can study independently with virtual tutors, and that educators can safely expect even students with poor academic skills to acquire a meaningful understanding of science. However, this may be true only if the virtual tutoring system is designed to sensitively address the needs of young learners—a conclusion that will be further supported by the detailed analysis of student responses discussed later on.

Was the MSB a useful and feasible educational tool in authentic educational settings?

Both students' self-reports and teachers' observations suggest that students' engagement was high while they were studying with the MSB. The student query answers indicated that the students liked the MSBs, felt they learned a lot, considered the content easy to learn, and wanted to study the MSBs often. The student query answers also showed that the MSBs increased the students' enthusiasm toward science. These results are in line with our assumption that highly structured lessons are well received by young children. Presumably, this is because highly structured lessons minimize the need to self-regulate the learning (Dignath and Büttner 2008), allowing the user to focus fully on the actual science content. In fact, some of the students appreciated the possibility of studying the MSBs without interruptions and with their headphones on, suggesting that young children enjoyed the independent working sessions—an observation also made by teachers.

The teachers appreciated the simple and easy-to-use interface of the program, and they mentioned that students with attention and self-regulation difficulties seemed to remain engaged while using the MSB. Teachers also expressed the need for scientific evidence on the efficiency of the MSBs in teaching science, and they were concerned about the pedagogical quality of the system. These are important aspects for any education technology that aims to be popular and effective. Thus, the virtual tutors are effective in teaching science, and they appear to be well received by students and teachers in the early grades.

Future work is needed to study how best to integrate such independent study sessions into classroom teaching. Given the promising learning and usability results, the MSB may serve as a tool for practicing independent studying while equipping all students, irrespective of their academic skills, with basic knowledge on a specific topic, which can then be elaborated in more collaborative learning settings. This may encourage students with poorer academic skills to participate in discussions about science.

How did the presence of an animated virtual tutor affect students' learning and/or enjoyment of the program in comparison to just hearing the tutor's voice?

Seeing the virtual tutor on screen (face and voice) did not produce larger learning gains than listening to the tutor's voice without her face on screen. While students provided positive feedback about the tutor, their views were split down the middle on whether they preferred that the tutor be visible. Four students reported that they did not like having the tutor on screen. This is in accordance with earlier studies suggesting that although virtual tutors are generally well received, some individuals do not like them (Gulz 2004; Schroeder and Adesope 2014). Future studies could provide students with the option of having the tutor on screen at any time within the application. We may learn, for example, that a significant proportion of students choose to remove the tutor's face during narrated science explanations but choose to have her face visible when answering MCQs.

What design features might have affected students' knowledge acquisition?

The error analysis suggested that in addition to common misconceptions, young students made errors due to perceptual confusions (Kameenui and Carnine 1998), exceptions to a rule, or the lack of an explicit rule on how to solve the science problems at hand.

The perceptual confusions are understandable given the young students' limited world knowledge. The lesson designers may want to either replace such challenging materials or address them in an intriguing way, thus providing opportunities for surprising and educational wrong answers. For example, the following prompt can be used when a student erroneously selects a frog instead of a fish as an animal covered with scales:

It's great that you selected a frog instead of a human or a bear. Remember that frogs have smooth skin, not scales, which are interlocked pieces of covering. You are right in the sense that this particular frog has coloring resembling very large scales. Can you come up with a picture of an animal with real scales?

Perhaps the most compelling example of prior conceptions influencing an answer choice was the question "Which animal has smooth skin?" Over a third of the students chose the human arm with hair on it even though the science presentation showed humans with hair on their bodies. It is likely that students (and many adults) believe that people have smooth skin (despite having hair on their skin), a belief that is reinforced by media advertisements for products that promote and show people with smooth skin. Such prior knowledge may be taken into account in the actual lessons:

You probably think that you have smooth skin. But if you would look at your hand through a magnifying glass, you would see small hairs growing all over. That's why we say that humans have hair. If you would look at a frog's skin, you wouldn't see even the tiniest hair. In science, we say that frogs have smooth skin and that humans have hair.

Elementary life science tends to teach the similarities and differences between animals (e.g., their skin coverings, ways of moving, life cycles, and unique characteristics) using simple examples, as was often the case in our multimedia lessons. Our error analysis suggested that providing an underlying principle whenever possible would be highly recommended. For example, students did not learn to differentiate very well crawling and walking, which were taught by representative examples. Instead, they learned to reliably identify insects from other creatures based on the number of body parts, the number of legs, and the presence of antennae.

Although the MSB were constructed based on extensive scientific knowledge on the effective means to teach science to small children through multimedia, the error analysis revealed several places for development. Some of the issues identified were almost impossible to foresee, such as the perceptual confusions. Therefore, future designers may need to engage in iterative development cycles for the multimedia presentations, in which the process of error analysis of the student responses, like the one presented here, seems to be crucial.

Limitations

This study did not include a control group because we were more interested in learning whether integrating MSB into classroom science instruction was feasible, whether the system could fully engage students, whether the system could increase students' interest in or excitement about science, and whether students learned the content that the lessons were designed to teach them. While the present study demonstrated that virtual tutors can teach meaningful science understanding in an engaging way to first-grade students, it did not compare the efficiency of virtual tutors as a teaching method with other possible methods such as typical classroom instruction. Thus, a clear next step is to improve the existing

multimedia lessons, as discussed above, and conduct a study comparing students in treatment and control conditions to learn whether students who use the MSB as a supplement to classroom instruction improve their science learning and become more motivated to study science in the future.

Conclusion

In this study, we explored the initial promise of a virtual tutoring system, the MSB, in engaging Finnish first-grade students in science learning. Results indicate that learners in the first year of school were already highly engaged and motivated by the program. They were consistently focused and “on task” while studying with MSBs, and the majority of students reported that they enjoyed the course and were more excited about science after it. Students answered about 89% of all questions correctly at the first attempt and answered 97% of the questions correctly at the second attempt, which was preceded by a hint. The pretest and posttest science assessment questions indicated significant learning gains on the taught material relative to learning gains on the not-taught content. Finally, these results did not differ regardless of whether the virtual tutor’s face was on or off screen when the tutor was speaking.

This study provides initial evidence that a sequence of narrated science explanations, followed by formative assessments that enable students to master core science concepts and vocabulary, can enhance students’ learning and provide teachers with useful tools for increasing their students’ interest in science, stimulating reasoning about science, and increasing their science knowledge. Further research needs to be conducted to examine the efficacy of the MSB in affecting learning compared with business-as-usual instruction.

Acknowledgements The authors would like to thank Sakke Hintsala and Maija Koikkalainen for their work in the data gathering phase.

Funding This study was part of the Science Across Virtual Institutions (SAVI; <http://www.innovationsforlearning.net/>) project funded by the US National Science Foundation (NSF) and the Academy of Finland. The research reported here was supported by the Academy of Finland through Grant 269102 to the Agora Center, University of Jyväskylä, Institute of Education Sciences, the U.S. Department of Education, through Grant R305B070008, and the National Science Foundation through Grants DRL 0733323 to Boulder Language Technologies, with a subcontract to Southern Methodist University, and through Grant DRL 0733322 to the University of Colorado at Boulder. The opinions expressed here are those of the authors and do not represent views of the Institute, the U.S. Department of Education, the National Science Foundation, or the Academy of Finland.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures in this study were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Appendix 1

See Table 3.

Table 3 Children's performance on pretest and posttest science assessment questions

#	A	Book	MCQ ACC %	Question	Gain Wilcoxon Z	Score pre	Score post	Correct answer	Most frequent wrong answer at posttest	Suggested explanation for no or poor learning gain, poor accuracy rate or frequent errors
1	-	1	98.4	What sense is needed for using a telescope?	- 1	0.9	0.95	Eyes	-	In ceiling
2	-	1	98.4	What sense do you need when someone uses a horn?	- 1.667*	0.85	0.93	Ears	-	In ceiling
3	-	1	98.4	What sense do you use for finding a light switch in the dark?	- 3.578***	0.54	0.8	Hand	Eyes 11%	Eyes also a reasonable answer
4	-	1	-	What senses do you need for enjoying a movie?	- 1.366	0.56	0.69	Eyes and ears	Mouth 24%	Combination of senses not taught; "Enjoying" may be associated with eating while watching a movie
5	T	4	87.1	Which animal crawls?	- 1.706*	0.34	0.48	Salamander	Snake 33%	Difference between slithering and crawling not taught at biodynamic level
6	-	4	100	Which animal uses wings and feet to move?	- 0.333	0.9	0.92	Dragonfly	-	In ceiling
7	-	4	61.3	Walks with four legs	- 0.218	0.7	0.69	Camel	Salamander 21%	Difference between walking and crawling not taught
8	T	5	90.3	Which pic represents a need for water?	- 4.491***	0.46	0.82	Horse drinking	-	
9	T	5	95.2	Which pic represents a need for shield?	- 3.638***	0.62	0.87	Rabbit at burrow	Dad and boy 8%	Dad may provide a shelter for his son
10	T	5	95.2	Which pic represents a need for oxygen?	- 2.524**	0.48	0.66	Dolphin at surface	Dad and boy 21%	Humans need also oxygen
11	-	5	77.4	Gets oxygen from water	- 2.496**	0.72	0.87	Fish or alike	Missing 11%	Open question, required writing
12	-	5	-	Separates oxygen from water	- 1.732*	0.08	0.13	Gills	Missing 70%	Not main focus of the book. No MCQ. Required writing

Table 3 continued

#	A	Book	MCQ ACC %	Question	Gain Wilcoxon Z	Score pre	Score post	Correct answer	Most frequent wrong answer at posttest	Suggested explanation for no or poor learning gain, poor accuracy rate or frequent errors
13	T	6	77.4	Which animal has scales?	- 2.840**	0.62	0.8	Fish	Frog 14%	Frog more rational error than human or bear. In addition, the frog had coloring resembling large scales
14	-	6	100	Which animal has fur?	- 1.633	0.97	0.9	Bear	-	In ceiling
15	T	6	85.5	Which animal has a pale skin?	- 1.528	0.48	0.59	Frog	Human 34%	No devoted MCQ that humans have hair
16	-	6	-	Which animal has hair	- 2.828**	0.11	0.31	Humans	Bear 48%	No devoted MCQ that humans have hair
17	-	7	85.5	Insect's bodypart for moving	- 2.646**	0.36	0.59	Thorax	Abdomen 19%	?
18	-	7	87.1	Insect's bodypart for digesting	- 3.772***	0.23	0.57	Abdomen	Thorax 24%	In humans the digestion is located at the center of the body
19	-	7	87.1	Insect's bodypart for sensing	- 1.147	0.72	0.8	Head	Abdomen 8%	?
20	-	7	93.1	Which is an insect?	- 5.100***	0.34	1.13	Dragonfly and bug	-	
21	T	8	79.0	Select a life cycle during a chrysalis forms	- 1	0.33	0.26	Caterpillar	Chrysalis 56%	Too difficult; some choose a pic of creature mentioned in the question
22	T	8	85.5	Select a life cycle that transforms into a butterfly	- 0.243	0.36	0.38	Chrysalis	Adult b-fly 45%	Too difficult; some choose a pic of creature mentioned in the question
23	T	8	85.5	Select a life cycle that hatches into a caterpillar	- 3.780***	0.23	0.56	Egg	Caterpillar 30%	Birth from egg easier to understand than later transformations. Some choose a pic of creature mentioned in the question
24	T	8	75.8	Select a life cycle while it breeds	- 1.800*	0.28	0.43	Butterfly	Egg 32%	Breeding involves not complex transformation. Birth from egg can be confused with breeding
25	N	NO		Which part of a plant produces oxygen?	- 1.4	0.28	0.39	Leaf		

Table 3 continued

#	A	Book	MCQ ACC %	Question	Gain Wilcoxon Z	Score pre	Score post	Correct answer	Most frequent wrong answer at posttest	Suggested explanation for no or poor learning gain, poor accuracy rate or frequent errors
26	N	NO		Which part of a plant delivers nutrients?	− 0.655	0.3	0.34	Roots		
27	N	NO		Which part of a plant produces seeds?	− 0.218	0.46	0.48	Flower		
28	N	NO		Which part of a plant transports nutrients?	− 1.46	0.3	0.41	Stalk		
29	N	NO		Which animal has a spine?	− 2.324**	0.62	0.77	Lamb		
30	N	NO		Which animal has an exterior structure?	− 0.894	0.67	0.74	Bug		
31	−	NO		Which animal has no supporting structure?	− 1.414	0.75	0.89	Medusa. Snail		
32	−	NO		Which animal gives birth?	0	0.52	0.52	Elk. Whale		
33	−	NO		Which animal lays eggs?	− 0.6	0.93	0.89	Eagle. Crocodile		
34	−	NO		Which animal spawns?	− 1.807*	0.25	0.36	Fish		
35	−	NO		Which animal is a mammal?	− 1.886*	0.46	0.33	Elk. Whale		
36	N	NO		Which body system distributes oxygen?	− 0.943	0.26	0.2	Circulation		
37	N	NO		Which body system is for ingestion	− 0.688	0.51	0.46	Digestion		
38	N	NO		Which body system supports us?	− 1.213	0.82	0.74	Skeleton		

Table 3 continued

#	A	Book	MCQ ACC %	Question	Gain Wilcoxon Z	Score pre	Score post	Correct answer	Most frequent wrong answer at posttest	Suggested explanation for no or poor learning gain, poor accuracy rate or frequent errors
39	N	NO		Which system guide/control us?	- 1.46	0.48	0.59	Nervous system		

T questions on taught content included in the analysis of variance, *N* questions on non-taught content included in the analysis of variance, *MCQ ACC* mean accuracy percentage on related multiple choice question in the MSBs

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.001$

Appendix 2

See Table 4.

Table 4 Children's MSBs MCQ error analysis

Book	Summary	MCQ accuracy	Most frequent errors	Possible explanation for errors
1. Five senses	Vision, hearing, touch, taste and smell were introduced with photo examples	96%	–	
2. How animals move?	Walking, crawling, slithering, flying and swimming were introduced including short descriptions of which body parts or limbs the animal uses for moving	90%	For the question “How do animals move?” the correct answer was “In many ways”, while 11/62 answered “Animals walk and run”, 10/62 answered “Animals use foot for moving” and 6/62 selected “Animals fly in the air”	These wrong answers are valid on their own if not taking account all choices, while the correct answer is a bit vague
3. What animals need?	Covered basic needs for water, food, oxygen, shelter and sleeping	89%	For the question “How do land animals get water?” the correct answer was “By drinking”, while 12/62 answered “By eating”	It was taught that animals get also some water from food
			For the question “What all animals need?” the correct answer was “All animals need space”, while 10/62 selected “All animals must live outside”	Some children may have understood that space equals with living outside—animals living indoor were not covered in the book
			For the question “What need this picture shows (a gecko having a dragonfly in its mouth)?” the correct answer was “Need for food”, while 8/62 selected “Need for oxygen”	These children may have understood that the dragonfly is dying due to lack of oxygen
4. How animals are covered?	Introduced fur (hair for humans), feathers, wet and dry scales, and amphibians pale skin	96%	For the question “Why are animals covered?” the correct answer was “For shield”, while 8/62 answered for “For locomotion” and 5/62 for “For hydration”	These answers are valid exceptions taught in the book

Table 4 continued

Book	Summary	MCQ accuracy	Most frequent errors	Possible explanation for errors
5. What is an insect?	This book was more specific than the previous ones, teaching that insects always have three pairs of legs, and three bodyparts (head, thorax and abdomen with their main functions) and antennae	85%	For the question “What are insect’s three main body parts?” the correct answer was “Head, thorax, abdomen” while 6/62 chose “Thorax, foot, and abdomen” or “Eyes, antenna, mouth”, and 5/62 chose “Head, abdomen, antenna” For the question “What organs all insects have in their head?” the correct answer was “Antenna, eyes, and jaws”, while 6/62 answered “Eyes, foot, wings” For the question “which body part is mainly for digestion” the correct answer was “Abdomen”, while 6/62 answered “Thorax”	Speculatively, this question was posed immediately after listing the main body parts, which potentially was insufficient to induce learning in some children This error is surprising as the alternative “Eyes, foot, wings” should have been the easiest one to rule out, due to containing two parts not attached to head In humans the digestion is located at the center of the body
6. Life cycle of a butterfly	Taught thoroughly egg, caterpillar, chrysalis and adult life cycles of a butterfly. The content of this book was more abstract by nature than the content of the other books	88%	In 7 out of 11 questions there was an erroneous alternative deriving 5 or more selections	These error responses seem to reflect the general difficulty of the book, instead of specific content of the lessons or MCQs

Appendix 3

Intervention assessment questionnaire

1. In my opinion, the computer books were: (Good, Between, Bad).
2. If you could choose, would you continue in reading the computer books in future? (Yes, No).
3. How often would you prefer to read these books? (More often, As now, Less often)
4. Which version of the books you prefer? (Talking head, No talking head)
5. Which way of studying you prefer? (In classroom, By computer)
6. Would you prefer to read the computer books: (At school, At home)
7. How excited are you about environmental science? (Less than previously, Same as previously, More than previously)

8. What did you like the most in the computer books?
9. What did you like the least in the computer books?

Teacher survey

1. According to your experiences, how useful MSB type of educational technology is in early education context?
2. How would you improve the educational value of such technology?
3. Is educational technology merely a burden, or more like a promise from the viewpoint of a teacher?
4. How would you evaluate the usability of the MSB for the children?
5. According to your perceptions, how much did children like the MSB?
6. In which setting you would prefer to use this kind of technology (e.g. in your classroom or in computer classroom)?
7. Other observations, notes and ideas for development?

References

- Adair, J. G., Sharpe, D., & Huynh, C. L. (1989). Hawthorne control procedures in educational experiments: A reconsideration of their use and effectiveness. *Review of Educational Research*, 59(2), 215–228.
- Adesope, O. O., & Nesbit, J. C. (2012). Verbal redundancy in multimedia learning environments: A meta-analysis. *Journal of Educational Psychology*, 104, 250–263. <https://doi.org/10.1037/a0026147>.
- Black, P., & William, P. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, 5(1), 7–73.
- Cole, R., Wise, B., & Van Vuuren, S. (2007). How Marni teaches students to read. *Educational Technology*, 47(1), 14.
- Craig, S., Gholson, B., Ventura, M., & Graesser, A. (2000). The tutoring research group: Overhearing dialogues and monologues in virtual tutoring sessions. *International Journal of Artificial Intelligence in Education*, 11, 242–253.
- Dalacosta, K., Kamariotaki-Paparrigopoulou, M., Palyvos, J. A., & Spyrellis, N. (2009). Multimedia application with animated cartoons for teaching science in elementary education. *Computers & Education*, 52(4), 741–748.
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231–264.
- Duit, R., & Treagust, D. F. (2003). Conceptual change: A powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671–688.
- Gardner, C. M., Simmons, P. E., & Simpson, R. E. (1992). The effects of CAI and hands-on activities on elementary students' attitude and weather knowledge. *School Science and Mathematics*, 92(6), 334–336.
- Gulz, A. (2004). Benefits of virtual characters in computer based learning environments: Claims and evidence. *International Journal of Artificial Intelligence in Education*, 14(3), 313–334.
- Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? *Educational Research Review*, 6(1), 27–54.
- Hong, S. Y., & Diamond, K. E. (2012). Two approaches to teaching young children science concepts, vocabulary, and scientific problem-solving skills. *Early Childhood Research Quarterly*, 27(2), 295–305.
- Kameenui, E. J., & Carnine, D. W. (1998). *Effective teaching strategies that accommodate diverse learners*. Des Moines: Prentice-Hall Inc.
- Kang, S., Scharmann, L. C., & Noh, T. (2004). Reexamining the role of cognitive conflict in science concept learning. *Research in Science Education*, 34(1), 71–96.
- Kintsch, E. (2005). Comprehension theory as a guide for the design of thoughtful questions. *Topics in Language Disorders*, 25(1), 51–64.
- Kupari, P., & Sulkunen, S. V. J. & Nissinen, K. (2012). *Enemmän iloa oppimiseen: Neljännen luokan oppilaiden lukutaito sekä matematiikan ja luonnontieteiden osaaminen. Kansainväliset PIRLS-ja*

- TIMSS-tutkimukset Suomessa. [More Joy into Learning: Reading, Math and Science proficiency in the Fourth Grade. International PIRLS- and TIMSS-studies in Finland.]* Helsinki: Ministry for Culture and Education, and University of Jyväskylä.
- La Paro, K. M., & Pianta, R. C. (2000). Predicting children's competence in the early school years: A meta-analytic review. *Review of Educational Research*, 70(4), 443–484.
- Lieberman, D. A., Bates, C. H., & So, J. (2009). Young children's learning with digital media. *Computers in the Schools*, 26(4), 271–283.
- Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., & Harrison, C. J. (2013). Animated agents and learning: Does the type of verbal feedback they provide matter? *Computers & Education*, 67, 239–249.
- Lindeman, J. (1998). *Allu-Ala-asteen lukutesti [Reading test for primary school]*. Turku, Finland: University of Turku.
- Mayer, R. E. (2014). *Cognitive theory of multimedia learning. The Cambridge handbook of multimedia learning*. New York: Cambridge University Press.
- National Center for Education Statistics. (2011). *The Nation's Report Card: Science 2009 (NCES 2011–451)*. Washington: Institute of Education Sciences.
- National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: National Academies Press.
- Osborne, J., & Dillon, J. (2008). *Science education in Europe: Critical reflections* (Vol. 13). London: The Nuffield Foundation.
- PISA, OECD. (2012). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know*. Retrieved from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf>
- Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners. *Journal of Research on Technology in Education*, 46(3), 229–251.
- She, H. C., & Liao, Y. W. (2010). Bridging scientific reasoning and conceptual change through adaptive web-based learning. *Journal of Research in Science Teaching*, 47(1), 91–119.
- Smith, J. P., III, Disessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences*, 3(2), 115–163.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.
- Van Lehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3), 227–265.
- Van Lehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221.
- Wang, F., Kinzie, M. B., McGuire, P., & Pan, E. (2010). Applying technology to inquiry-based learning in early childhood education. *Early Childhood Education Journal*, 37(5), 381–389.
- Ward, W., Cole, R., Bolanos, D., Buchenroth-Martin, C., Svirsky, E., Vuuren, S. V., et al. (2011). My science tutor: A conversational multimedia virtual tutor for elementary school science. *ACM Transactions on Speech and Language Processing (TSLP)*, 7(4), 18.
- Ward, W., Cole, R., Bolaños, D., Buchenroth-Martin, C., Svirsky, E., & Weston, T. (2013). My science tutor: A conversational multimedia virtual tutor. *Journal of Educational Psychology*, 105(4), 1115.
- Wise, B., Cole, R., Van Vuuren, S., Schwartz, S., Snyder, L., Ngampatipatpong, N., et al. (2005). Learning to read with a virtual tutor: Foundations to literacy. In C. Kinzer & L. Verhoeven (Eds.), *Interactive literacy education: Facilitating literacy environments through technology* (pp. 31–75). Mahwah, NJ: Lawrence Erlbaum.
- Yoshida, R. (2008). Teachers' choice and learners' preference of corrective feedback types. *Language Awareness*, 17(1), 78–93.

Jarkko Hautala is a Post-Doctoral Researcher in the Center for Applied Language Studies and Department of Psychology at the University of Jyväskylä, Finland. His research interests include reading, eye-tracking, learning disabilities and science learning.

Doris Luft Baker is an Assistant Professor in the Department of Teaching & Learning at Southern Methodist University, Dallas, Texas. Her research interests include vocabulary and science learning, bilingual assessments to guide instruction, and parent engagement.

Aleksi Keurulainen worked as a programmer and researcher at the Agora Center, University of Jyväskylä, Finland.

Miia Ronimus is Post-Doctoral Researcher at the Niilo Mäki Institute, Jyväskylä, Finland. Her research interests include engagement in technology-enriched learning, and dyslexia.

Ulla Richardson is a Professor in Centre for Applied Language Studies at the University of Jyväskylä, Finland. Her research interests include development of computer games for learning to read, and early language development.

Ronald Cole is the Principal Scientist & President of Boulder Learning, Colorado. His research interest is to develop educational technologies for language and science learning.